

# Broad-UNet: Multi-scale feature learning for nowcasting tasks

Jesús García Fernández, Siamak Mehrkanoon\*

Department of Data Science and Knowledge Engineering, Maastricht University, The Netherlands



## ARTICLE INFO

### Article history:

Received 10 February 2021

Received in revised form 21 May 2021

Accepted 30 August 2021

Available online 10 September 2021

### Keywords:

Satellite imagery

Precipitation forecasting

Cloud cover forecasting

Deep learning

Convolutional neural network

U-net

## ABSTRACT

Weather nowcasting consists of predicting meteorological components in the short term at high spatial resolutions. Due to its influence in many human activities, accurate nowcasting has recently gained plenty of attention. In this paper, we treat the nowcasting problem as an image-to-image translation problem using satellite imagery. We introduce Broad-UNet, a novel architecture based on the core UNet model, to efficiently address this problem. In particular, the proposed Broad-UNet is equipped with asymmetric parallel convolutions as well as Atrous Spatial Pyramid Pooling (ASPP) module. In this way, the Broad-UNet model learns more complex patterns by combining multi-scale features while using fewer parameters than the core UNet model. The proposed model is applied on two different nowcasting tasks, i.e. precipitation maps and cloud cover nowcasting. The obtained numerical results show that the introduced Broad-UNet model performs more accurate predictions compared to the other examined architectures.

© 2021 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Weather forecasting is an essential task that has a great influence on humans daily life and activities. Industries such as agriculture (Cogato, Meggio, De Antoni Migliorati, & Marinello, 2019), mining (Ivanov, Ivanova, & Kuvshinkin, 2019) and construction (Senouci, Al-Abbasi, & Eldin, 2018) rely on the weather forecasts to make decisions and thus unexpected climatological events may result in large economic losses. Similarly, accurate weather forecasts improve safety on flights and roads and help us foresee potential natural disasters.

Due to its importance, precipitation nowcasting is becoming an increasingly popular research topic. This term refers to the problem of forecasting precipitation in the near future at high spatial resolutions. It is usually performed through satellite imagery and many different approaches have been proposed for this problem. Classical nowcasting approaches mainly focus on two methods: Numerical Weather Prediction (NWP) (Sun et al., 2014) and extrapolation based techniques, such as Optical Flow (OF) (Woo & Wong, 2017). NWP methods simulate the underlying physics of the atmosphere and ocean to generate predictions, so they require a vast amount of computational resources. In contrast, optical flow based methods identify and predict how objects move through a sequence of images. But they are unable to represent the dynamics behind them. In recent years, the massive amount of existing data has aroused research

interest in data driven machine learning techniques for nowcasting (Grover, Kapoor, & Horvitz, 2015; Holmstrom, Liu, & Vo, 2016; Shi et al., 2015). By taking advantage of available historical data, data-driven based approaches have shown better performance than classical ones in many forecasting tasks (Faloutsos, Gasthaus, Januschowski, & Wang, 2019). Furthermore, while classical machine learning techniques rely on handcrafted features and domain knowledge, deep learning techniques automatize the extraction of those features. Recent advances in deep learning have shown promising results in diverse research areas such as neuroscience, biomedical signal analysis, weather forecasting and dynamical systems, among others (Alaoui Abdellaoui, García Fernández, Şahinli, & Mehrkanoon, 2021; Coban, 2013; Coban & Aksu, 2018; Gamboa, 2017; Mehrkanoon, 2019a, 2019b; Mehrkanoon & Suykens, 2015, 2018; Salman, Kanigoro, & Heryadi, 2015; Webb, 2018). Convolutional Neural Networks (CNNs) are the most popular algorithms used in computer vision (Voulodimos, Doulamis, Doulamis, & Protopapadakis, 2018), achieving the state-of-the-art in various tasks (Goel, Sharma, & Kapoor, 2020; Lu & Weng, 2007; Voulodimos et al., 2018). CNN architectures, such as AlexNet (Krizhevsky, Sutskever, & Hinton, 2017), ResNet (He, Zhang, Ren, & Sun, 2016) and Inception-Net (Szegedy et al., 2015), to name a few, mainly consist of the combination of convolutional and pooling layers. They are outstanding at classification, identification and recognition tasks. Among other architectures, autoencoders have emerged as one of the most powerful approaches in both supervised (Berthomier, Pradel, & Perez, 2020; Fernández, Abdellaoui, & Mehrkanoon, 2020; Zhang et al., 2019) and unsupervised learning (Baldi, 2012; Chung, Wu, Shen, Lee, & Lee, 2016; Lample, Conneau, Denoyer,

\* Corresponding author.

E-mail addresses: [j.garciafernandez@student.maastrichtuniversity.nl](mailto:j.garciafernandez@student.maastrichtuniversity.nl) (J.G. Fernández), [siamak.mehrkanoon@maastrichtuniversity.nl](mailto:siamak.mehrkanoon@maastrichtuniversity.nl) (S. Mehrkanoon).

& Ranzato, 2017) with the UNet (Ronneberger, Fischer, & Brox, 2015) being one of the most versatile architectures. The UNet architecture was first proposed for medical image segmentation, but it has been employed in various domains (Fernández et al., 2020; Tao, Palasek, Ling, & Patras, 2017; Trebing, Stańczyk, & Mehrkanoon, 2021). It consists of a contracting path, to extract features, and an expanding path, to reconstruct a segmented image, with a set of residual connection between them to enable precise localization. In our previous work (Fernández et al., 2020), we introduced various extended versions of the UNet for weather forecasting problem. In this paper, we further extend the best performing model in that work (Fernández et al., 2020), i.e. the AsymmlIncepRes3DDR-UNet. In particular, motivated by the results of Chen, Papandreou, Kokkinos, Murphy, and Yuille (2017), we augment the AsymmlIncepRes3DDR-UNet's feature extraction capacity by incorporating an Atrous Spatial Pyramid Pooling module (ASPP) (Chen et al., 2017) in the bottleneck of the network. The ASPP module works in line with the existing building blocks of our network (Multi-scale feature convolutional block), extracting multi-scale features in parallel and combining them. Therefore, unlike the original UNet, the proposed model is designed to capture multi-scale information. In addition, it keeps the temporal dimension unchanged along the encoder path and then reduces it before being concatenated with the output of every level in the decoder path. As a result, it can efficiently learn a mapping between 3-dimensional input data and 2-dimensional output data. Furthermore, we apply a kernel factorization in most of the convolutional operations of the model, resulting in a significant reduction in the total number of parameters compared to the original UNet while having improved performance. These techniques are explained in detail in the subsequent sections. We further present an analysis of this multi-scale features extraction and the enhancement provided by the ASPP module. We show its versatility by applying it to two different nowcasting tasks, i.e. precipitation nowcasting and cloud cover nowcasting. In the precipitation nowcasting task, the model performs a regression of every pixel. In the case of cloud cover nowcasting, the model classifies each pixel as containing clouds or not. In addition, we directly compare the proposed model with the model introduced in Trebing et al. (2021), a variation of UNet architecture that relies on depthwise-separable convolutions and includes a CBAM attention module (Woo, Park, Lee, & Kweon, 2018) at each level. While the model in Trebing et al. (2021) approximates the performance of the original UNet with a significantly reduced number of parameters, our model outperforms the original UNet with a reduced number of parameters.

## 2. Related work

Traditionally, optical flow based models are the most popular techniques among classical methods for precipitation nowcasting tasks (Bowler, Pierce, & Seed, 2004; Li et al., 2018). However, machine learning and deep learning based approaches are dominating this field of research in recent years. Due to the vast amount of available satellite imagery, powerful deep neural networks based models are suitable candidates that can be used to address various problems existing in this field. In particular, CNN based architectures have shown their great ability to handle 2D and 3D images. Thanks to the versatility of CNN's, nowcasting problems can be tackled in different fashions. For instance, the authors in Agrawal et al. (2019) and Ayzel, Heistermann, Sorokin, Nikitin, and Lukyanova (2019) treated the multiple time-steps as multiple channels in the network. In this way, they could apply a simple 2D-CNN to perform the predictions. Additionally, the authors in Shi et al. (2017) treated the multiple time-steps as depth in the samples. Thus they can apply a 3D-CNN and approximate more complex functions. As it has been shown in Agrawal

et al. (2019), Lebedev et al. (2019) and Trebing et al. (2021), among the used CNN architectures, UNet is more suitable for this task, due to its autoencoder-like architecture and ability to tackle image-to-image translation problems.

In addition to CNN's, Recurrent Neural Networks (RNN's) have proved to be a robust approach. However, these architectures struggle to work with images but can capture long-range dependencies, an ability that CNN's can only partially achieve with the addition of attention mechanisms, such as self-attention (Vaswani et al., 2017). In Shi et al. (2015), the authors introduce an architecture that combines both CNN's and RNN's strengths. They extend the fully connected LSTM (FC-LSTM) with convolutional structures, obtaining the Convolutional LSTM network (ConvLSTM). As a result, the proposed model captures spatiotemporal correlations better than the FC-LSTM model. The authors in Shi et al. (2017) introduce the Trajectory GRU (TrajGRU) model as an extension of the ConvLSTM. This architecture keeps the advantages of the previous model and also learns the location-variant structure of the recurrent connections, showing superior performance than the other models compared. Nevertheless, these RNN models have not been directly compared with the UNet in nowcasting tasks.

The authors in Berthomier et al. (2020) make a comparison among different types of models for cloud cover nowcasting. In Berthomier et al. (2020), the models under assessment are various versions of CNN's, RNN's, LSTM and UNet. The authors showed that the UNet model is the best performing model for the given cloud cover nowcasting task.

## 3. Proposed model

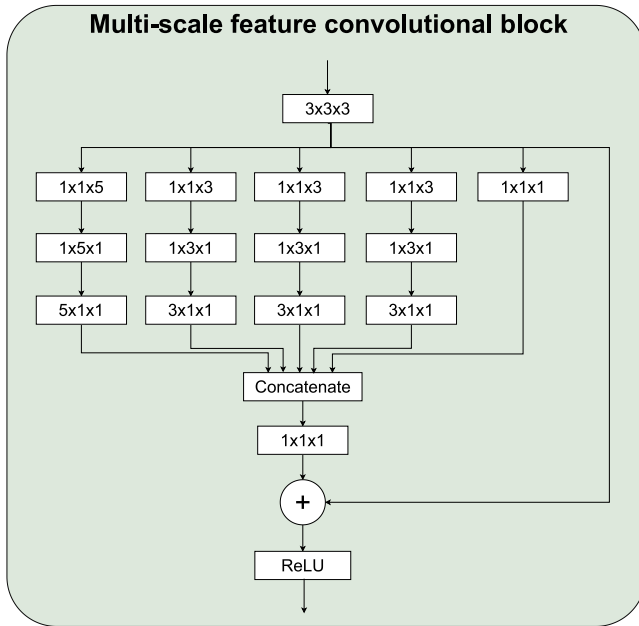
In this section, we introduce our Broad-UNet model. First, different elements that are used for building the network are presented. The complete architecture is then explained.

### 3.1. Multi-scale feature convolutional block

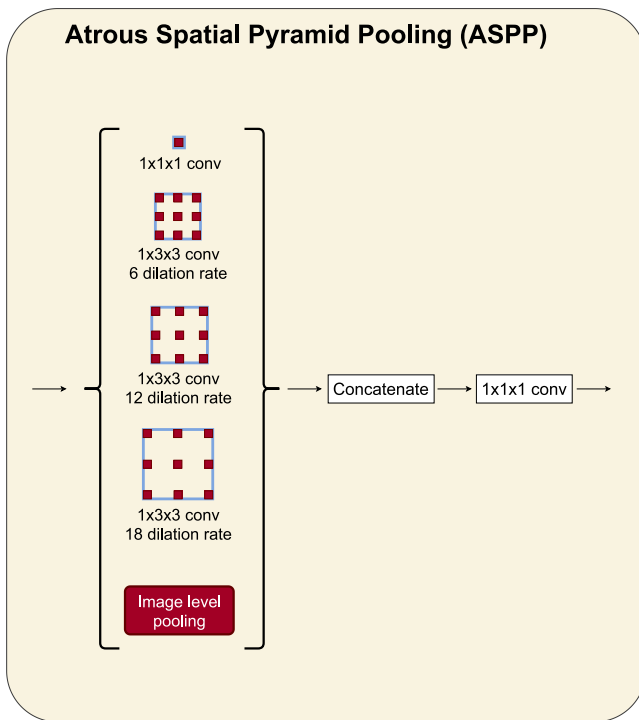
Motivated by the goal of extracting features at different scales, the model contains a block consisting of parallel arms as shown in Fig. 1. This block serves as the core building block of our network. Within this block, the data forks into parallel branches of convolutions with different kernel sizes, after going through an initial convolution. A  $3 \times 3 \times 3$  convolution is followed by a set of parallel convolutions with  $1 \times 1 \times 1$ ,  $3 \times 3 \times 3$  and  $5 \times 5 \times 5$  kernel sizes. The outputs of the different branches are then concatenated and merged with a  $1 \times 1 \times 1$  convolution. Additionally, inspired by the results found in Szegedy, Ioffe, Vanhoucke, and Alemi (2016), we keep some information intact alongside the parallel branches with a residual connection. Lastly, the output of the block is rectified with a ReLU activation function. To reduce the large number of features resulting from these branches, we factorize the convolutions as suggested in Yang et al. (2019). That means a convolution  $N \times N \times N$  decomposes into the three consecutive  $1 \times 1 \times N$ ,  $1 \times N \times 1$  and  $N \times 1 \times 1$  convolutions. Hence, this sequence is an approximation of the original convolution with fewer parameters.

### 3.2. Atrous Spatial Pyramid Pooling (ASPP)

Atrous Spatial Pyramid Pooling (ASPP), is a mechanism used to capture multi-scale information. It consists of parallel branches of convolutions, similar to the convolutional block presented above. However, instead of using different kernel sizes, the same kernel is chosen with an increasing dilation rate (6, 12 and 18). In this kind of convolutions, the filter is upsampled by inserting zeros between successive values. As a result, they employ a larger field



**Fig. 1.** Multi-scale feature convolutional block. Convolutions with different kernels are performed in parallel to extract features at different scales. A residual connection also keeps some unmodified information.



**Fig. 2.** Atrous Spatial Pyramid Pooling (ASPP) block. Different dilation rates allow the network to extract multi-scale information. Due to the kernel shapes ( $1 \times N \times N$ ) and the image-level pooling mechanism, only spatial information is extracted.

of view, without experiencing an explosion in the number of parameters. Further, it only extracts information in the spatial dimensions by applying a 2-dimensional filter (shape  $1 \times N \times N$ ). In addition, ASPP incorporates one branch to extract image-levels features, allowing to capture global context information.

**Table 1**

Comparison between the number of learnable parameters in different UNet based models examined in this paper.

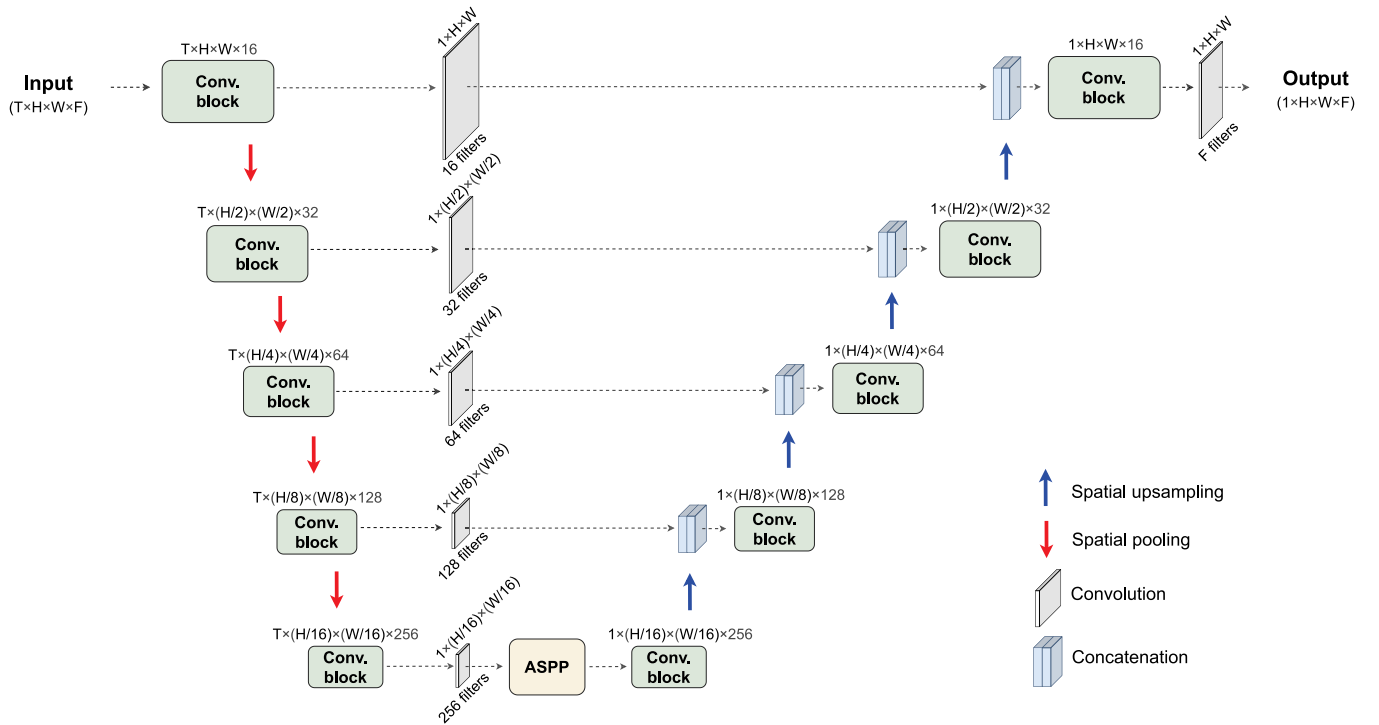
Model	Number of parameters
UNet	~17M
SmaAt-UNet	~4M
AsymmlIncepRes3DDR-UNet	~9.5M
Broad-UNet	~11M

Here, we implement it by applying a global average pooling, and subsequent reshaping and upsampling back. The obtained extracted features are then concatenated and combined with a  $1 \times 1 \times 1$  convolution. The scheme of this mechanism is shown in Fig. 2.

### 3.3. Broad-UNet

Thanks to the effectiveness of UNet architecture in solving image-to-image mapping tasks, it is chosen to serve as the basis to construct our model. UNet core model which was originally proposed for medical image segmentation tasks, adopts an auto-encoder structure. While the encoder part extracts features from the input image, the decoder part performs classification on each pixel to reconstruct the segmented output. Plus, a set of residual connections between both parts allows a precise localization in the output image. Differently, our proposed Broad-UNet manipulates 3-dimensional data in the encoder and 2-dimensional data in the decoder. Thus, we can input several time-steps in the first dimension, and it outputs only one time-step in the same dimension. Multi-scale feature convolutional blocks are alternated with pooling operations in the encoder, resulting in five levels. The pooling is only performed in the spatial dimensions (2nd and 3rd) and implemented with a Max Pooling layer. In this way, the temporal dimension of the data remains unchanged. Then, the decoder follows a similar structure. It alternates multi-scale feature convolutional blocks and upsampling in the spatial dimensions. Additionally, we incorporate extra convolutions in the connections between different levels of the encoder and decoder. These intermediate convolutional operations aim to reduce the temporal dimension from  $T$  time-steps to 1.

To extend the multi-scale feature learning process, we combine the convolutional blocks with the ASPP module. It is placed in the bottleneck of the network, where the data has a highly abstract representation. In this way, we allow the network to capture more information from this representation without using larger kernels and more computational resources. Also, dropout is included in the bottleneck to force the network to learn a more sparse representation of the data and avoid possible overfitting. As a result, the network input is of shape  $T \times H \times W \times F$  and output is of shape  $1 \times H \times W \times F$ , where  $T$  is the number of time-steps (lags),  $H$  and  $W$  are the height and width of the images, and  $F$  is the number of features or elements, which we consider as channels in our network. Here, the convolutions to reduce the temporal dimension have a kernel size  $1 \times T \times T$  with valid padding. In addition, the use of asymmetric convolutions drastically reduces the total number of parameters of the network. While the number of parameters is ~28 million using regular kernels  $N \times N \times N$ , the number of parameters after factorizing the convolutions into  $1 \times 1 \times N$ ,  $1 \times N \times 1$  and  $N \times 1 \times 1$  is ~11 million. The complete architecture of the model can be found in Fig. 3. Furthermore, a comparison in the number of learnable parameters among different UNet based models examined in this paper is shown in Table 1.



**Fig. 3.** Complete architecture of Broad-UNet. The multi-scale feature convolutional block is displayed as *Conv. Block* for simplicity. The annotation over these blocks describes the output dimension, where  $T$  represents the time-steps (lags),  $H$  and  $W$  represent the height and width of each image and  $F$  the number of features or elements predicting.

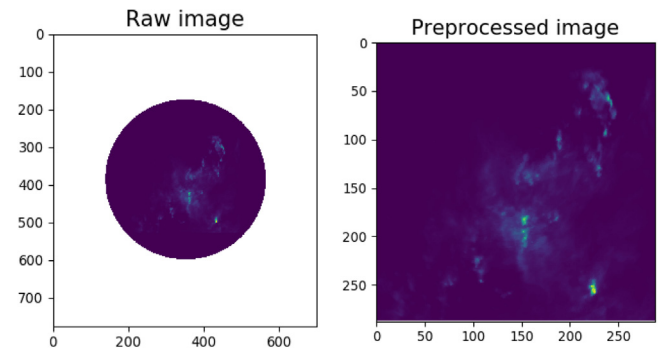
#### 4. Data description and preprocessing

To assess the performance of our model, we apply it to two different datasets. Both of them consist of satellite images and are intended to tackle weather nowcasting problems. The first one includes precipitation maps, in which the value of each pixel shows the amount of rainfall in that region. The second one consists of cloud cover maps, in which the pixel values are binary and indicate whether there is a cloud or not in that region. Here, we recreate the same samples as in Trebing et al. (2021) for the first dataset, and as in Berthomier et al. (2020) for the second dataset. In this way, we can make a fair comparison with the results obtained in those research works. For reproducibility purposes, all our models and scripts are available on Github.<sup>1</sup> Also, the datasets and pre-trained models are available upon request.

##### 4.1. Precipitation maps dataset

The first dataset, provided by the Royal Netherlands Meteorological Institute (Koninklijk Nederlands Meteorologisch Instituut, KNMI) (Institute, 2020), includes rainfall measurements from two Dutch radar stations (De Bilt and Den Helder). These measurements are in the shape of images. The images cover the region of the Netherlands and neighbouring countries, spanning four years in 5-min intervals. To train and validate the models, we use data from the years 2016–2018 (80% train/20% validation), and the data from 2019 is used as test set.

The values of each pixel represent the accumulated amount of rainfall in the last five minutes. That means that a value  $n$  represents  $n \times 10^{-2}$  mm of rain per square kilometre. The resolution of the images is  $765 \times 700$ , and the measured region is circle-shaped with a large margin. Following the lines of Trebing et al. (2021), we cropped the central squared area with size  $288 \times 288$ , as shown in Fig. 4.



**Fig. 4.** Example of precipitation map from the first dataset, before and after applying the preprocessing.

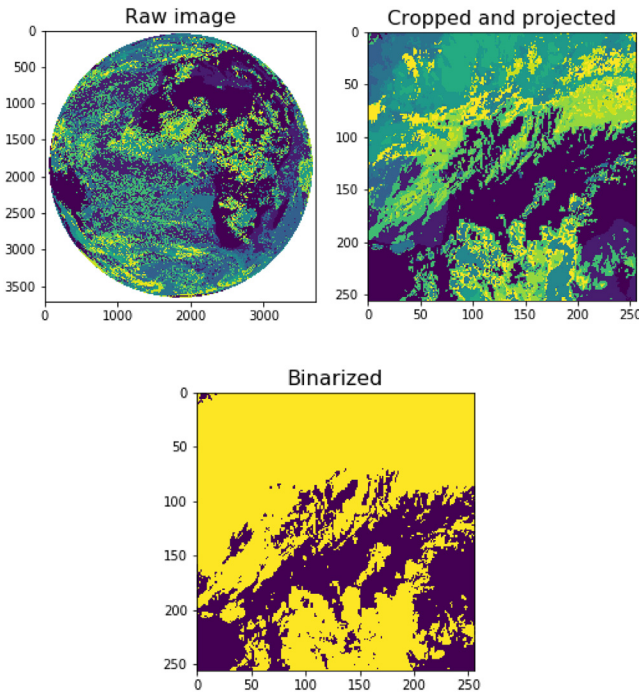
Moreover, there is a high imbalance between pixels with rain and no rain, with plenty of images lacking raining pixels. Therefore, as in Trebing et al. (2021), we filter the dataset choosing only the images with at least 50% of pixels containing any amount of rain. This dataset is then used to create the training/validation/test samples. Additionally, we create a second dataset filtering the images with at least 20% of pixels containing any amount of rain. From this second dataset, we use only the test set. Therefore, it serves as a way of testing our trained models under different conditions. We also normalize both datasets by dividing them by the highest value in the training set.

##### 4.2. Cloud cover dataset

The second dataset is the “Geostationary Nowcasting Cloud Type” classification product (EUMETSAT, 2020) from the European Organisation for the Exploitation of Meteorological Satellites (EUMETSTAT). It is composed of satellite images of the whole

<sup>1</sup> <https://github.com/jesusgf96/Broad-UNet>.





**Fig. 5.** Example of an image from the cloud cover dataset, before and after applying the different steps of the preprocessing.

globe at longitude 0 degrees, taken every 15 min. The resulting size of the images in the dataset is  $3712 \times 3712$ , spanning the years 2017–2018. For multi-comparison purposes, we generate two different dataset from this data. In the first dataset, we follow the lines of Berthomier et al. (2020) and use data from 2017 and the first semester of 2018 as training set. Then the data from the second semester of 2018 is used for both validation and test. On the contrary, as in Trebing et al. (2021), we use data from 2017 and the first semester of 2018 for train and validate our models (80% train/20% validation). The data from the second semester of 2018 is thus used only for test. We use data from 2017 and the first semester of 2018 to train the models. To validate and test the models, we use data from the second semester of 2018. In this data, every pixel can have 15 different values (1: Cloud-free land, 2: Cloud-free sea, 3: Snow over land, 4: Sea ice, 5: Very low clouds, 6: Low clouds, 7: Mid-level clouds, 8: High opaque clouds, 9: Very high opaque clouds, 10: Fractional clouds, 11: High semitransparent thin clouds, 12: High semitransparent meanly thick clouds, 13: High semitransparent thick clouds, 14: High semitransparent above low or medium clouds, 15: High semitransparent above snow/ice). However, following the lines of Berthomier et al. (2020), we aim to perform a classification between cloud or no-cloud. Therefore, we group the labels from 1 to 4 into 0 (no-cloud) and the labels from 5 to 15 into 1 (cloud). Also, we crop the images according to the boundaries of France: [51.896, 41.104, -5.842, 9.842] (upper latitude, lower latitude, left longitude, right longitude). Then we apply a transformation to obtain a suitable projection and reshape the resulting image to  $256 \times 256$  pixels. Fig. 5, displays an example of the described pre-processing steps.

## 5. Experimental setup and evaluation

In order to have a fair comparison with the results obtained in Berthomier et al. (2020) and Trebing et al. (2021) for both datasets, we reproduce the same experimental setups. The data

is arranged in such a way that the resulting input is a four-dimensional array  $\mathcal{I} \in \mathbb{R}^{T \times H \times W \times F}$ , where  $T$  is the number of lags or previous time-steps, which corresponds to the time dimension.  $H$  and  $W$  refer to the size of the image and make up the spatial dimensions. The last element  $F$  corresponds to the predicted features, which in both cases is 1. We use TensorFlow to implement our models and train and evaluate them on the given datasets. The hyperparameters of models are tuned and the optimal ones are empirically found and used.

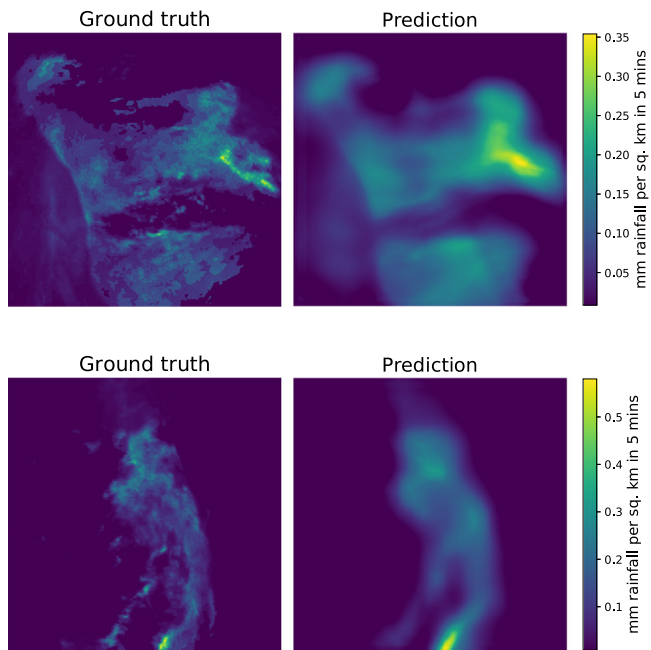
### 5.1. Precipitation maps nowcasting

As for the precipitation maps dataset, we apply the preprocessing and split the dataset as described in Section 4.1. We aim to predict a precipitation map 30 min ahead or considering that the images are generated five minutes apart, six time-steps ahead. The number of lags, previous time-steps, is set to 12 which was empirically found to be the best one among other tested lag values. The height and width of the images are 288 and 288, and the number of features in the input is 1, i.e. the precipitation maps. Therefore, the inputs of the model has the shape (12, 288, 288, 1), and output data has the shape (1, 288, 288, 1).

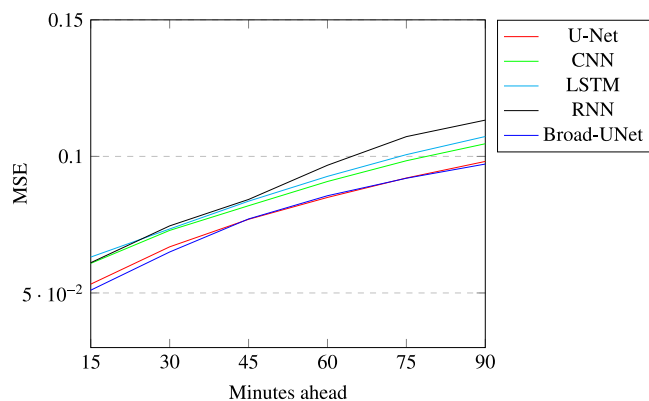
In this nowcasting task, we perform a regression of every pixel. Mean Squared Error (MSE) is used as the loss function and Adam optimizer to optimize it, with an initial learning rate of 0.0001. The batch size and the dropout rate are set to 2 and 0.5, respectively. We also implemented a checkpoint callback to monitor the validation loss. Thus the best performing model on the validation set is saved. We use MSE as the main metric to assess the performance of the model in this dataset. Furthermore, we also include additional metrics such as accuracy, precision and recall. Following the lines of Trebing et al. (2021), in order to calculate these new metrics, we first create a binarized mask of the image, according to a threshold. This threshold is the mean value of the training set from the 50% of rain pixel dataset. Hence, any value equal or over the threshold is replaced by 1, and any value under it is replaced by 0.

### 5.2. Cloud cover nowcasting

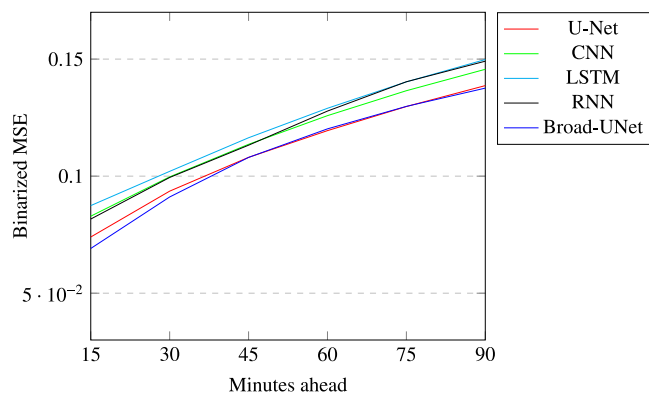
Regarding the cloud cover dataset, we preprocess the data and split the dataset as described in Section 4.2. In this case, we predict six different time-steps: from 15 min to 70 min ahead, or from 1 to 6 time-steps ahead. Due to the architecture of our network, we train six different model. Thus each model predicts a different time-step. Here, the number of lag is set to 4, the height and width of the images are 256 and 256, and the number of input features is again 1, the cloud cover map. That means that the model receives input data with the shape (4, 256, 256, 1), and outputs data with the shape (1, 256, 256, 1). In this task, we perform binary classification of every pixel. The binary cross-entropy is used as the loss function in this case. We use Adam optimizer with an initial learning rate of 0.001. The batch size and the dropout rate are set to 8 and 0.5, respectively. Similarly, we implemented a checkpoint callback to monitor the validation loss. Thus the best performing model on the validation set is saved. Following the lines of Berthomier et al. (2020), here we also use MSE as the metric to assess the performance of the model. First, we calculate the MSE between the ground truth and the raw prediction as the main metric. In this case, the values between 0 and 1 in the predictions indicate the probability of cloud occurrence in that region. In addition, we binarize the prediction of the network with a threshold of 0.5 to generate a second assessment with the MSE metric. We also include additional metrics, i.e. accuracy, precision and recall, to compare the performance of the Broad-UNet with the model introduced in Trebing et al. (2021), which also uses UNet architecture as the basis. To calculate these new metrics, we first create a binarized mask of the image, using the value 0.5 as the threshold.



**Fig. 6.** Broad-UNet precipitation prediction examples. The images in the first row are generated with the test set from the 50% of pixels containing rain dataset. The images in the second row are generated with the test set from the 20% of pixels containing rain dataset.



**Fig. 7.** Test MSE values of the different models for the cloud cover prediction task.



**Fig. 8.** Test MSE values of the different models for the cloud cover prediction task with binarized predictions.

**Table 2**

Test MSE and additional metrics values for the precipitation maps prediction task using the 50% of rain pixels dataset. ↓ indicates that the optimal values are the smallest ones and ↑ indicates that the optimal values are the highest ones.

Model	MSE 50% of rain pixels dataset			
	MSE ↓	Accuracy ↑	Precision ↑	Recall ↑
Persistence	2.48e−02	0.756	0.678	0.643
UNet	1.22e−02	0.836	0.740	0.855
SmaAt-UNet	1.22e−02	0.829	0.730	0.850
AsymmIncepRes3DDR-UNet	1.11e−02	0.858	0.759	0.800
Broad-UNet	1.08e−02	0.850	0.715	0.817

**Table 3**

Test MSE and additional metrics values for the precipitation maps prediction task using the 20% of rain pixels dataset. ↓ indicates that the optimal values are the smallest ones and ↑ indicates that the optimal values are the highest ones.

Model	MSE 20% of rain pixels dataset			
	MSE ↓	Accuracy ↑	Precision ↑	Recall ↑
Persistence	2.28e−02	0.827	0.559	0.543
UNet	1.11e−02	0.880	0.666	0.782
SmaAt-UNet	1.11e−02	0.867	0.626	0.801
AsymmIncepRes3DDR-UNet	1.02e−02	0.893	0.621	0.767
Broad-UNet	1.02e−02	0.895	0.611	0.772

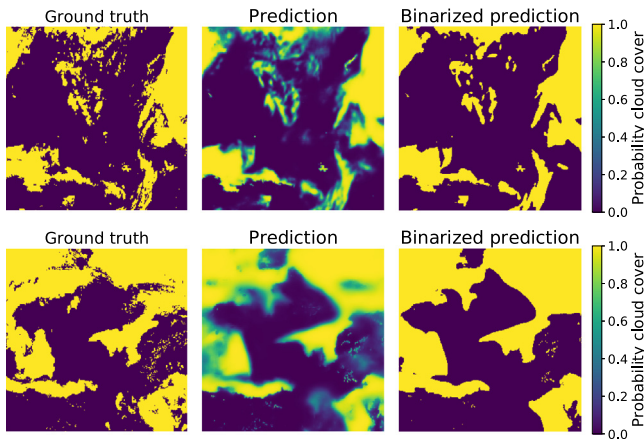
## 6. Results

### 6.1. Precipitation maps nowcasting

In the precipitation maps prediction task, we compare the performance of the Broad-UNet with the persistence, a simple meteorological baseline used in forecasting, and different models over the test sets of two different datasets, i.e. 50% of rain pixels and 20% of rain pixels. These models are the UNet (Ronneberger et al., 2015) and two variants (Fernández et al., 2020; Trebing et al., 2021). The MSE is the main metric used for this comparison, and it is calculated over the denormalized data. The additional metrics are computed over the binarized data, as described in Section 5.1. The performance of different models over the first precipitation maps dataset is shown in Table 2. In the same way, the performance of the models in the second precipitation maps dataset is listed in Table 3. From the obtained results, one can observe that the Broad-UNet achieved the lowest MSE score in both datasets. Two examples of 30 min ahead prediction with the Broad-UNet are displayed in Fig. 6. The images on the first and second row are generated using the first and second precipitation dataset, respectively.

### 6.2. Cloud cover nowcasting

When applying the Broad-UNet to the second dataset, we compare its performance with the persistence and various models. These models are introduced and explained in Berthomier et al. (2020). We perform this comparison with the results obtained from the test set of the cloud cover dataset. The used evaluation metrics are explained in Section 5.2. In Fig. 7, we show the MSE obtained using the ground truth and the actual prediction. Fig. 8 depicts the MSE calculated with the ground truth and binarized prediction. From Figs. 7 and 8, one can notice that the Broad-UNet performance is superior in short-term forecasting. As the number of step-ahead increases, the gap between the performance of the proposed Broad-UNet and the classical UNet model decreases. In addition, in Table 4, we show the comparison between the Broad-UNet's and the model introduced in Trebing



**Fig. 9.** Broad-UNet cloud cover prediction examples. The image above is predicted 30 min ahead, and the image below is predicted 90 min ahead.

**Table 4**

Average of the MSE and additional metrics for the cloud cover prediction task. ↓ indicates that the optimal values are the smallest ones and ↑ indicates that the optimal values are the highest ones.

Model	Averaged test MSE cloud cover prediction			
	MSE ↓	Accuracy ↑	Precision ↑	Recall ↑
Persistence	0.1491	0.851	0.849	0.849
UNet	0.0785	0.890	0.895	0.919
SmaAt-UNet	0.0794	0.889	0.892	<u>0.921</u>
Broad-UNet	<u>0.0783</u>	<u>0.891</u>	<u>0.898</u>	0.914

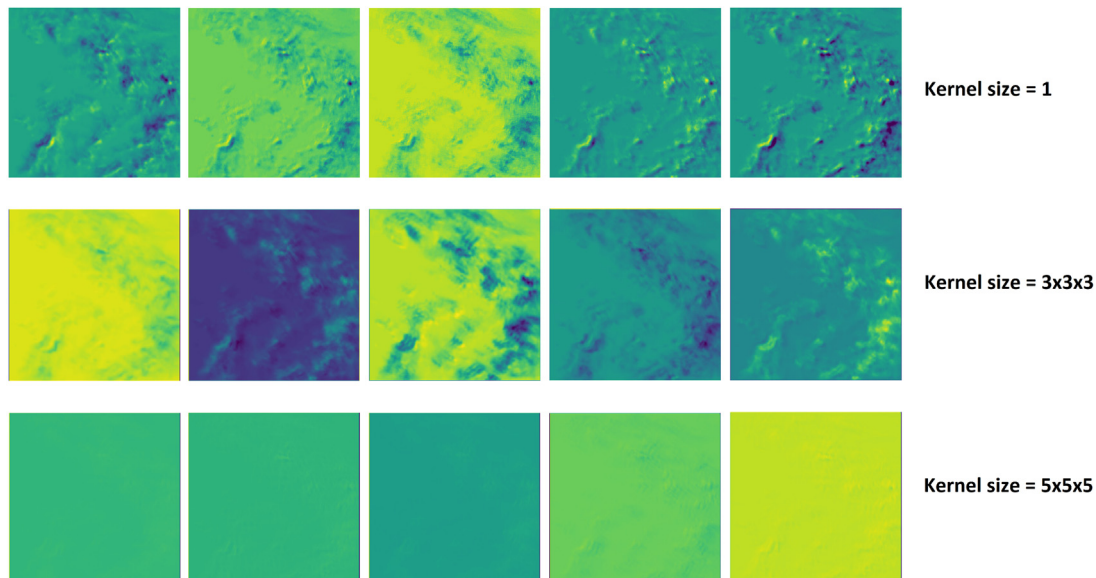
et al. (2021). As in Trebing et al. (2021), the metrics tabulated in this table are averaged over different time-steps (15–90 min ahead). From the obtained results one can observe that the Broad-UNet performs better than other compared models in three out of four used metrics. In addition, two examples of the Broad-UNet's predictions are displayed in Fig. 9. Both predictions are generated with the test set of the cloud cover dataset. In Fig. 9, the images on the first and second row correspond to 30 min and 90 min ahead prediction, respectively.

## 7. Discussion

From the obtained results, one can observe that the multi-scale feature learning allows the Broad-UNet to perform more precise predictions. This is thanks to the use of different convolutional filters in parallel. By combining convolutions with larger and smaller kernels, the model considers different amounts of information around the same region to generate the feature maps. Likewise, the inclusion of the ASPP module in the architecture allows the network to apply convolutions with diverse receptive fields at the same time.

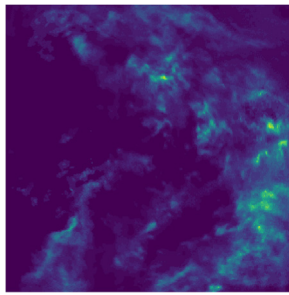
In the precipitation nowcasting task, we can observe an 11% and an 8% improvement with respect to the simple UNet for both datasets. In the cloud cover nowcasting task, the binarized predictions of the Broad-UNet are 7% more accurate than the simple UNet for 15 min ahead predictions, and 1% more accurate for 90 min ahead predictions. Since in the first nowcasting task, i.e. precipitation prediction, the model aims to perform a regression of each pixel with a wide range of values, achieving accurate forecasting or equivalently lower MSE values is more desirable. That is where the Broad-UNet shows more superior performance respect to the UNet. In the second nowcasting task, where the goal is to carry out a binary classification on each pixel, Broad-UNet performs slightly more accurate predictions than the UNet. While the immediate predictions (i.e. 15 and 30 min ahead) are more precise, more distant predictions (more than 45 min ahead) are comparable to UNet's predictions. Therefore, we can state that the wide building blocks of the Broad-UNet let the network to extract the spatial and short-term temporal information more accurately than the regular UNet.

The learnt feature maps in different branches inside a convolutional block is shown in Fig. 10. The chosen convolutional block is the first one so that the data does not have too abstract representation and is thus easier to interpret. The image fed to the network belongs to the precipitation maps dataset, and it is shown in Fig. 11. In Fig. 10, the first row of the feature maps is the output of the convolutional branch with kernel size 1. The second row corresponds to the output of the branch with kernel size  $3 \times 3 \times 3$ . Lastly, the third row corresponds to the output of the branch with kernel size  $5 \times 5 \times 5$ . From Fig. 11, one can observe the differences between the features extracted in each



**Fig. 10.** Feature maps outputted by different branches inside the first multi-scale feature convolutional block. Every row represents the output of a different branch. Next to each row, the kernel employed by the convolution in that branch is shown.





**Fig. 11.** Image fed to the network to generate the feature maps shown below. It belongs to the precipitation maps datasets, specifically to the 50% of rain pixel dataset.

branch. The convolutions with kernel size 1 seem to strengthen detailed differences in the image, and convolutions with kernel size  $3 \times 3 \times 3$  seem to accentuate differences between areas containing a high and a low rain concentration. In addition, convolutions with kernel size  $5 \times 5 \times 5$  seem to highlight regions with high rain concentration.

## 8. Conclusion

In this paper the Broad-UNet, an extension of the UNet architecture, is introduced for precipitation as well as cloud cover nowcasting. Thanks to the combination of the multi-scale feature convolutional block and the incorporation of the ASPP module, the proposed network is able to capture multi-scale information. In addition, the use of factorized kernels drastically reduces the number of parameters in the network compared to classical UNet model. The performance of Broad-UNet is examined for addressing two nowcasting problems. The first problem consists of predicting precipitation maps 30 min ahead. The second one consists of forecasting cloud cover 15 to 90 min ahead. The obtained results suggest that the Broad-UNet extracts features more efficiently and therefore performs more accurate predictions in short-term nowcasting tasks compared to other tested UNet based models.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgement

We would like to thank Léa Berthomier from Météo France for providing the cloud cover dataset, as well as the required code to preprocess it.

## References

Agrawal, S., Barrington, L., Bromberg, C., Burge, J., Gazen, C., & Hickey, J. (2019). Machine learning for precipitation nowcasting from radar images. *arXiv preprint arXiv:1912.12132*.

Alaoui Abdellaoui, I., García Fernández, J., Şahinli, C., & Mehrkanoon, S. (2021). Enhancing brain decoding using attention augmented deep neural networks. Accepted for publication in the Proceedings of the 29th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, Oct, 2021..

Ayzel, G., Heistermann, M., Sorokin, A., Nikitin, O., & Lukyanova, O. (2019). All convolutional neural networks for radar-based precipitation nowcasting. *Procedia Computer Science*, 150, 186–192.

Baldi, P. Autoencoders, unsupervised learning, and deep architectures. In *Proceedings of ICML workshop on unsupervised and transfer learning* (pp. 37–49).

Berthomier, L., Pradel, B., & Perez, L. (2020). Cloud cover nowcasting with deep learning. *arXiv preprint arXiv:2009.11577*.

Bowler, N. E., Pierce, C. E., & Seed, A. (2004). Development of a precipitation nowcasting algorithm based upon optical flow techniques. *Journal of Hydrology*, 288(1–2), 74–91.

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2017). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4), 834–848.

Chung, Y.-A., Wu, C.-C., Shen, C.-H., Lee, H.-Y., & Lee, L.-S. (2016). Audio word2vec: Unsupervised learning of audio segment representations using sequence-to-sequence autoencoder. *arXiv preprint arXiv:1603.00982*.

Coban, R. (2013). A context layered locally recurrent neural network for dynamic system identification. *Engineering Applications of Artificial Intelligence*, 26(1), 241–250.

Coban, R., & Aksu, I. O. (2018). Neuro-controller design by using the multifeedback layer neural network and the particle swarm optimization. *Tehnički Vjesnik*, 25(2), 437–444.

Cogato, A., Meggio, F., De Antoni Migliorati, M., & Marinello, F. (2019). Extreme weather events in agriculture: A systematic review. *Sustainability*, 11(9), 2547.

EUMETSAT (2020). Geostationary nowcasting cloud type - MSG - 0 degree. URL <https://navigator.eumetsat.int/product/EO:EUM:DAT:MSG:GNWCCT>.

Faloutsos, C., Gasthaus, J., Januschowski, T., & Wang, Y. (2019). Classical and contemporary approaches to big time series forecasting. In *Proceedings of the 2019 international conference on management of data* (pp. 2042–2047).

Fernández, J. G., Abdellaoui, I. A., & Mehrkanoon, S. (2020). Deep coastal sea elements forecasting using U-Net based models. *arXiv preprint arXiv:2011.03303*.

Gamboa, J. C. B. (2017). Deep learning for time-series analysis. *arXiv preprint arXiv:1701.01887*.

Goel, R., Sharma, A., & Kapoor, R. (2020). State-of-the-art object recognition techniques: A comparative study. In *Soft computing: Theories and applications* (pp. 925–932). Springer.

Grover, A., Kapoor, A., & Horvitz, E. (2015). A deep hybrid model for weather forecasting. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 379–386).

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770–778).

Holmstrom, M., Liu, D., & Vo, C. (2016). Machine learning applied to weather forecasting. *Stanford University*, 2–4.

Royal Netherlands Meteorological Institute (2020). Royal netherlands meteorological institute. URL <https://www.knmi.nl/home>.

Ivanov, S., Ivanova, P., & Kuvshinkin, S. (2019). Weather conditions as a factor affecting the performance of modern powerful mining excavators. In *Journal of physics: Conference series*, Vol. 1399. IOP Publishing, Article 044070.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), 84–90.

Lample, G., Conneau, A., Denoyer, L., & Ranzato, M. (2017). Unsupervised machine translation using monolingual corpora only. *arXiv preprint arXiv:1711.00043*.

Lebedev, V., Ivashkin, V., Rudenko, I., Ganshin, A., Molchanov, A., Ovcharenko, S., et al. (2019). Precipitation nowcasting with satellite imagery. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining* (pp. 2680–2688).

Li, L., He, Z., Chen, S., Mai, X., Zhang, A., Hu, B., et al. (2018). Subpixel-based precipitation nowcasting with the pyramid Lucas-Kanade optical flow technique. *Atmosphere*, 9(7), 260.

Lu, D., & Weng, Q. (2007). A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing*, 28(5), 823–870.

Mehrkanoon, S. (2019a). Cross-domain neural-kernel networks. *Pattern Recognition Letters*, 125, 474–480.

Mehrkanoon, S. (2019b). Deep neural-kernel blocks. *Neural Networks*, 116, 46–55.

Mehrkanoon, S., & Suykens, J. A. K. (2015). Learning solutions to partial differential equations using LS-SVM. *Neurocomputing*, 159, 105–116.

Mehrkanoon, S., & Suykens, J. A. K. (2018). Deep hybrid neural-kernel networks using random Fourier features. *Neurocomputing*, 298, 46–54.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International conference on medical image computing and computer-assisted intervention* (pp. 234–241). Springer.

Salman, A. G., Kanigoro, B., & Heryadi, Y. (2015). Weather forecasting using deep learning techniques. In *2015 international conference on advanced computer science and information systems (ICACSIS)* (pp. 281–285). IEEE.

Senouci, A., Al-Abbasi, M., & Eldin, N. N. (2018). Impact of weather conditions on construction labour productivity in Qatar. *Middle East Journal of Management*, 5(1), 34–49.

Shi, X., Chen, Z., Wang, H., Yeung, D.-Y., Wong, W.-K., & Woo, W.-c. (2015). Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Advances in Neural Information Processing Systems*, 28, 802–810.



- Shi, X., Gao, Z., Lausen, L., Wang, H., Yeung, D.-Y., Wong, W.-k., et al. (2017). Deep learning for precipitation nowcasting: A benchmark and a new model. In *Advances in neural information processing systems* (pp. 5617–5627).
- Sun, J., Xue, M., Wilson, J. W., Zawadzki, I., Ballard, S. P., Onvlee-Hooimeyer, J., et al. (2014). Use of NWP for nowcasting convective precipitation: Recent progress and challenges. *Bulletin of the American Meteorological Society*, 95(3), 409–426.
- Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. (2016). Inception-v4, inception-resnet and the impact of residual connections on learning. arXiv preprint [arXiv:1602.07261](https://arxiv.org/abs/1602.07261).
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., et al. (2015). Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1–9).
- Tao, Y., Palasek, P., Ling, Z., & Patras, I. (2017). Background modelling based on generative unet. In *2017 14th IEEE international conference on advanced video and signal based surveillance (AVSS)* (pp. 1–6). IEEE.
- Trebing, K., Stańczyk, T., & Mehrkanoon, S. (2021). SmaAt-UNet: Precipitation nowcasting using a small attention-UNet architecture. *Pattern Recognition Letters*, 145, 178–186.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). Attention is all you need. In *Advances in neural information processing systems* (pp. 5998–6008).
- Voulodimos, A., Doulamis, N., Doulamis, A., & Protopapadakis, E. (2018). Deep learning for computer vision: A brief review. *Computational Intelligence and Neuroscience*, 2018.
- Webb, S. (2018). Deep learning for biology. *Nature*, 554(7693).
- Woo, S., Park, J., Lee, J.-Y., & Kweon, I. S. (2018). Cbam: Convolutional block attention module. In *Proceedings of the European conference on computer vision (ECCV)* (pp. 3–19).
- Woo, W.-c., & Wong, W.-k. (2017). Operational application of optical flow techniques to radar-based rainfall nowcasting. *Atmosphere*, 8(3), 48.
- Yang, H., Yuan, C., Li, B., Du, Y., Xing, J., Hu, W., et al. (2019). Asymmetric 3d convolutional neural networks for action recognition. *Pattern Recognition*, 85, 1–12.
- Zhang, J., Xie, Y., Zhang, P., Chen, H., Xia, Y., & Shen, C. (2019). Light-weight hybrid convolutional network for liver tumor segmentation. In *IJCAI* (pp. 4271–4277).